Using genetic and epigenetic data to predict human facial features – developing a theoretical basis for face prediction.

Human physical appearance is highly heritable and very distinctive by variation in facial characteristics, pigmentation, height, and body mass index. Elucidating the genetic basis of modern human diversity is an important issue of the contemporary human biology. Today, the diversity of modern humans can be effectively studied using genomic tools.

The main objective of the proposed project is to use genomic information to develop a theoretical basis for the prediction of facial appearance that could be useful in forensic and anthropological contexts. An additional aim is to generate data including rare genetic variants that could be combined in a collaborative study to further elucidate the genetics of facial morphology. Most of the data needed for the project has already been collected and requires bioinformatic and statistical analysis. Additional data will be collected using a dual facial phenotype analysis system, i.e. postmortem computer tomography and 3-dimensional scanning. The dual method of phenotyping will provide insight into the role of genes in shaping the hard (skeletal) and soft tissues of the face. This new sample set will be analysed using whole genome sequencing and Infinium MethylationEPIC microarrays, which provides DNA methylation information. This omics dataset will capture the influence of genes and environment on facial shape.

The highly complex phenotype of the face allows for mutual recognition and is therefore applicable to human identification. In addition, facial appearance is aesthetically important while there are many genetic disorders that cause facial malformations, of which cleft lip and palate is one of the most common birth defects. Therefore, the study of facial genetics is important in the context of research into the human evolution, forensic science, and for medical and social reasons. Despite the number of studies on facial genetics, they are still ongoing due to the need to collect large amounts of data, particularly to elucidate the importance of rare genetic variants and to clarify the epigenetic factor in shaping the three-dimensional facial phenotype. In this grant application, we propose to investigate the importance of DNA variation (both genetic and epigenetic) in explaining the facial phenotype. From the data available to date, a significant amount of information about facial phenotype is contained in genetic ancestry, as well as sex, age and BMI. Our research is expected to enable the prediction of facial appearance through the development of a mathematical algorithm (a prediction pipeline). The use of machine learning methods will be an essential part of the project, both in the phenotype description phase and in the development of the prediction method.

In addition, the data generated will allow its integration into the wider research stream of facial genetics and further elucidation of the complex genetic architecture of facial morphology.